



OPEN

Phenotypic characterization and genome analysis of a novel *Salmonella* Typhimurium phage having unique tail fiber genes

Sadia Sattar^{1,8}, Inam Ullah^{1,8}, Sofia Khanum², Marc Bailie², Bushra Shamsi¹, Ibrar Ahmed³, Syed Tahir Abbas Shah⁴, Sundus Javed⁵, Aamir Ghafoor⁶, Amna Pervaiz¹, Fakiha Sohail¹, Naseer Ali Shah¹, Kaleem Imdad¹, Nazish Bostan¹ & Eric Altermann^{2,7}

Salmonella enterica serovar Typhimurium is a foodborne pathogen causing occasional outbreaks of enteric infections in humans. *Salmonella* has one of the largest pools of temperate phages in its genome that possess evolutionary significance for pathogen. In this study, we characterized a novel temperate phage *Salmonella* phage BIS20 (BIS20) with unique tail fiber genes. It belongs to the subfamily *Peduvirinae* genus *Eganvirus* and infects *Salmonella* Typhimurium strain (SE-BS17; Acc. NO MZ503545) of poultry origin. Phage BIS20 was viable only at biological pH and temperature ranges (pH7 and 37 °C). Despite being temperate BIS20 significantly slowed down the growth of host strain for 24 h as compared to control ($P < 0.009$). Phage BIS20 features 29,477-base pair (bp) linear DNA genome with 53% GC content and encodes for 37 putative ORFs. These ORFs have mosaic arrangement as indicated by its ORF similarity to various phages and prophages in NCBI. Genome analysis indicates its similarity to *Salmonella enterica* serovar Senftenberg prophage (SEStP) sequence (Nucleotide similarity 87.7%) and *Escherichia virus* 186 (~82.4% nucleotide similarity). Capsid genes were conserved however those associated with tail fiber formation and assembly were unique to all members of genus *Eganvirus*. We found strong evidence of recombination hotspot in tail fiber gene. Our study identifies BIS20 as a new species of genus *Eganvirus* temperate phages as its maximum nucleotide similarity is 82.4% with any phage in NCBI. Our findings may contribute to understanding of origin of new temperate phages.

Salmonella Typhimurium, a major foodborne pathogen with zoonotic potential is responsible for several outbreaks of salmonellosis in humans resulting in high levels of morbidity and economic losses. Most outbreaks are due to the consumption of contaminated livestock and food products^{1,2}. The prevalence of these serovars in the food chain leads to high chances of human infections³. It is important to study the factors responsible for host adaptability in different ecological niches. *Salmonella* harbors a wide variety of temperate phages in their genomes. The role of temperate phages is critical for host adaptability during infection and for disease epidemiology. Particularly, phages as mobile genetic elements are found to be a source of genetic differences between avirulent and virulent strains of *E. coli* and other members of *Enterobacteriaceae*⁴. Prophages integrated into the host genome can encode genes that are not required for phage production instead, they can be of benefit to the host cell. These genes can enhance the bacterial host's fitness either by producing toxins or by increasing the fitness of the host during infection⁵. Organism tracing during a disease outbreak utilizes phage typing and serovar identification to distinguish isolates of the same serovar⁶. For example, recent publications from the United Kingdom reported the presence of a specific prophage in an invasive strain of *Salmonella* Typhimurium

¹Molecular Virology Labs, Department of Biosciences, Comsats University Islamabad, Islamabad, Pakistan. ²AgResearch, Palmerston North, New Zealand. ³Alpha Genomics Private Limited, Islamabad 45710, Pakistan. ⁴Functional Genomics Lab, Department of Biosciences, Comsats University Islamabad, Islamabad, Pakistan. ⁵Microbiology and Immunology Lab, Department of Biosciences, Comsats University Islamabad, Islamabad, Pakistan. ⁶University Diagnostic Lab, University of Veterinary and Animal Sciences (UVAS) Lahore, Lahore, Pakistan. ⁷Riddet Institute, Massey University, Palmerston North, New Zealand. ⁸These authors contributed equally: Sadia Sattar and Inam Ullah. ✉email: sadia.sattar@comsats.edu.pk

responsible for causing salmonellosis, that was not reported before in any UK-based strain^{7–9}. Hence temperate phage integration in host genomes exerts a directional pressure on host evolution and adaptability.

Most dsDNA-tailed temperate phages have mosaic genomes originating from different evolutionary backgrounds with interchangeable modules¹⁰. This mosaic module arrangement implies that mosaic genomes have originated by horizontal gene transfer of different functional modules from a common ancestor^{9,11,12}. Mosaic structure is best studied in lambdoid and T4-like bacteriophages^{13,14}. Phage genome regions associated with tail fibers exhibit a high degree of mosaicism and are recombination hot spots¹⁵. These recombination events may be the result of host/ environmental selection pressure on phages for finding new hosts or to adapt to an already evolved host, however, temperate phages may be subjected to completely different types of pressures as faced by lytic phages. Integrated prophages will undergo the same selection pressure as endured by host chromosome. In fact, even defective prophages can confer beneficial phenotypes to the host^{16,17}, conversely these defective prophages can also provide functional genes to integrated temperate phages that result in new phage variants capable of infecting new hosts (host range expansion). There is a constant exchange of genes between integrated prophages and temperate phages by homologous recombination that makes the evolutionary pattern more versatile¹⁸. In this study we isolated and characterized a novel *Salmonella* temperate phage BIS20 that exhibits mosaic genome arrangement and has sequence similarity to *Salmonella* prophage SEStP and *Escherichia* phage 186. However, BIS20 differs from it in repressor system and tail fiber genes. We found a clear evidence of recombination hotspot in tail fiber gene of BIS20 that may contribute to its evolutionary significance by shedding light into the processes that lead to host range changes among bacteriophages. BIS20 may represent an intermediate link between *Salmonella* prophage SEStP and *Escherichia* phage 186.

Results

Bacteriophage isolation and characterization. *Salmonella* phage BIS20 was isolated from retail poultry tissue samples suspected of *Salmonella* infection using *Salmonella* Typhimurium strain SE-BS17 (Acc. # MZ503545) of poultry origin. BIS20 produced milky/turbid zones of lysis on SE-BS17 (clearing zone 0.9 mm) a characteristic that suggests it is a temperate phage later confirmed by presence of integration cassette in its genome (Fig. 1A)^{19–21}. Host adsorption efficiency was 96% in 5 min after addition of BIS20 in SE-BS17 mid log phase culture (OD 0.35). The phage latent period as determined by a one-step growth curve, was 20 min with a burst size of 110 ± 7 PFU/cell (Fig. 1B, Supplementary Table S2). BIS20 host range was tested on 14 *Salmonella* strains. These strains were characterized up to subspecies level (materials and methods) and belonged to *Salmonella enterica* subspecies *enterica* however their serovars were not known. BIS20 produced clear zones of lysis on five strains only. Out of these five strains it grew best on SE-BS17, a *Salmonella* Typhimurium isolate as indicated by 16S rRNA sequencing (Supplementary Table S3). No lysis zones were produced on a strain of *S. Typhi* tested. Few other species of family *Enterobacteriaceae* tested by spot method were also not lysed. It is difficult to conclude the definitive host range of BIS20 as fewer strains were available for testing.

BIS20 produced maximum titer at 37 °C (1.7×10^6); however, it was still able to infect at temperatures up to 45 °C (1.9×10^2), albeit with a significant drop in titers ($P < 0.001$). BIS20 lost its infectivity after incubation at temperatures above 45 °C (e.g., 60 °C, 80 °C and 100 °C) as no plaques were observed at these temperatures (Fig. 1C). Highest titer of BIS20 was observed at pH7 ($\sim 1.72 \times 10^5$ PFU/ml), whereas a significant ($P < 0.001$) drop in titer was observed at pH 9 (1.23×10^1) after 1 h of incubation (Fig. 1D). No plaques / clearing zones were observed at pH 2, 5 and 12. Despite its temperate nature BIS20 was able to restrict growth of SE-BS17 up to 18 h post incubation at MOI 1 {BIS20 = 4.07×10^6 , control (SE-BS17) = 6.18×10^7 } ($P < 0.0001$). The maximum growth of the SE-BS17 strain observed after 24 h of incubation with BIS20 at MOI 1 was 1.3×10^7 CFU/ml which was significantly lower than bacteria only control (7.78×10^7 CFU/ml) ($P < 0.009$) (Fig. 1E). Transmission electron microscope analysis indicated that BIS20 has a hexagonal head with a tail (Fig. 2).

Genome characterization. *Salmonella* phage BIS20 has a 29,477-base pair (bp) genome and encodes for 37 Open Reading Frames (ORFs). The genome map of BIS20 is given in Fig. 3. All ORFs are color coded according to their putative functions. The BIS20 genome has 53% GC content, well in alignment with its host strain (53.2%). Thirty-four ORFs started with an ATG start codon, one ORF coding for a replication endonuclease protein started with TTG (ORF 32), and two other ORFs started with GTG that include DNA damage inducible protein I (Din I) responsible for initiating SOS response (ORF 33), and a conserved hypothetical protein (ORF 1). The phage has a modular organization; each module is labelled by roman numerals in Fig. 3. Modules I and III include ORFs that encode structural proteins (ORFs colored red and black). Module I include ORFs involved in capsid formation, head stabilization and DNA translocation to empty capsids (Fig. 3, ORFs 1–7). These highly conserved proteins have homology with P2 family of *Enterobacteriaceae* temperate bacteriophages. ORFs in third module were involved in phage tail, tail sheath, tail fibers and base plate assembly (Fig. 3, ORFs 11–24). Majority of these proteins were conserved except for tail fiber domain containing protein (ORF 17) and tail fiber assembly protein (ORF 18). These two proteins of BIS20 have high homology with SEStP only (Supplementary Table S5) but not with phages. In addition, phage late control gene D protein, involved in lysis of cell and virion maturation (ORF24) was also present in this module. Module II includes host cell lysis enzymes and regulatory proteins (ORFs 7–10 colored green). BIS20 utilizes the usual conserved Lambda phage holin, lysozyme: spannin complex for bacterial wall lysis. All three components; holin, endolysin (lysozyme) and spannin show high gene and protein homology with *Salmonella enterica* and *Citrobacter* sp. prophages. These proteins carried putative conserved protein domains of *Escherichia* phage P2 lysis cassette. In addition, BIS20 features a GP25-like lysozyme, likely a structural component of base plate outer wedge with acidic lysozyme activity involved in phage entry into the host cell by binding with peptidoglycan (ORF14). Module IV has ORFs involved in phage integration (tyrosine recombinase (Xer C) /integrase) and lytic to lysogenic switching (*CI*; immunity repres-

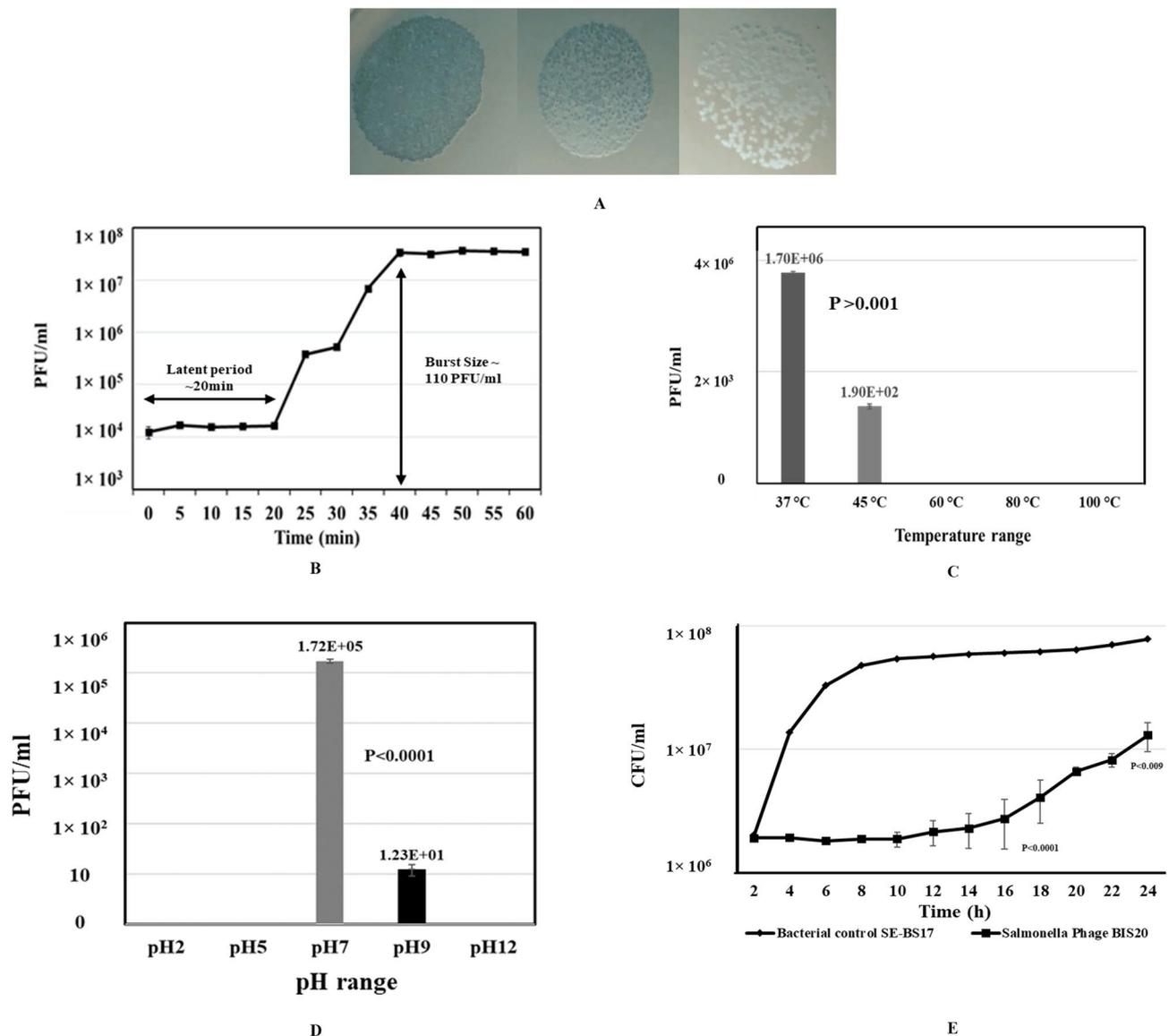


Figure 1. Physical Characterization of *Salmonella* phage BIS20: (A) BIS20 plaque morphology, turbid zones of lysis indicative of temperate nature of phage (B) Latent period and burst size of BIS20 determined by one step growth curve method (C) Temperature stability of BIS20; Phage was stable at 37 °C, barely survived at 45 °C, no plaques were observed at other temperatures tested. (D) BIS20 pH stability between 2 and 12 pH range; BIS20 grew best at pH7 and barely survived at pH 9 (E) Bacterial growth reduction assay; reading was monitored for 24 h, control; *Salmonella* Typhimurium (SE-BS17) in LB; SE-BS17 bacterial culture infected with *Salmonella* phage BIS20 at MOI1. All values are represented as average of triplicates.

tor, *ApI*, *excionase* and *CII*; *Lysogeny regulation*) (ORFs 25–28; colored orange). Module V includes hypothetical proteins (ORFs colored blue). Two hypothetical proteins (ORF 35 and 36) were unique to BIS20. Detailed description of each ORF and their putative function with homology is given in Supplementary Tables S4 and S5.

BIS20 is a linear dsDNA phage as no contigs assembled across both ends of the phage genome. Phage genome termini were analyzed by Phage Term software that utilizes two methods, a phage term method and Li's method²². Analysis indicated presence of multiple termini on forward and reverse strands. Peaks localized 20 bases around the maximum indicated presence of preferred termini with terminal redundancy and apparition of partially circular permutations. The phage genome has direct terminal repeats or terminal redundancies. Such arrangement of termini protects linear DNA phages from DNA losses during replication. The packaging of such DNA takes place in to already made proheads that requires use of ATP dependent terminase activity to sort DNA concatemers (Supplementary Fig. S2).

Phylogenetic analysis. *BIS20 similarity with prophages.* When BIS20 genome was compared in BLASTn the highest homology was shown by prophage sequences however genome coverage was low. As per ICTV recommendation VIRIDIC software was used to compare BIS20 genome with close phage and prophage homologs in BLASTn. VIRIDIC software is a tool to identify intergenomic similarities of bacteriophages and allow their

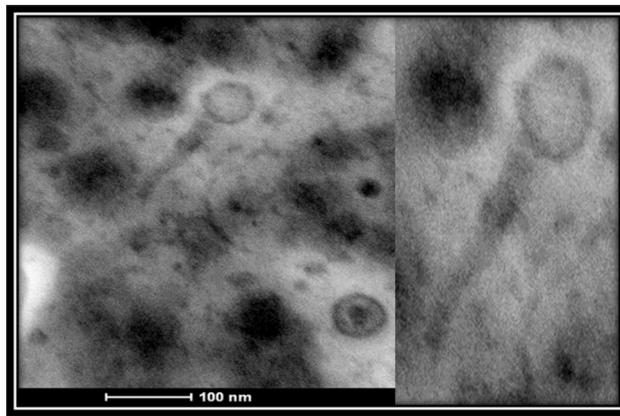


Figure 2. Transmission Electron Micrograph of *Salmonella* phage BIS20. Phage was stained with 1% uranyl acetate solution images were taken at an acceleration voltage of 80 kV. Scale bar represents 100 nm.

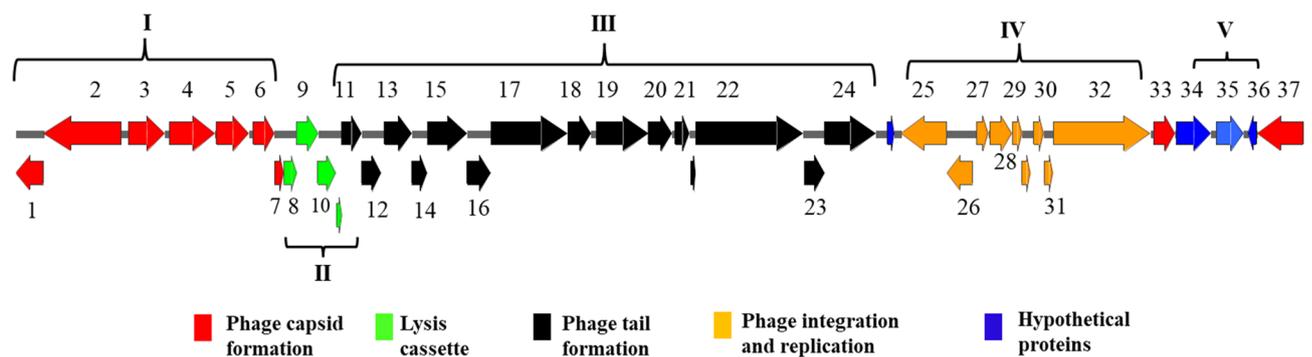


Figure 3. *Salmonella* phage BIS20 annotation map as generated by GeneMark.hmm, RAST and GAMOLA version 2 software. Different ORFs are color coded according to their putative functions. Numbers at each ORF correspond to their description in Supplementary Tables S4 and S5. For ease of description genome is divided in to five modules represented by Roman numerals and parentheses.

classification into a genus or a separate species. According to VIRIDIC the BIS20 phage genome has highest homology of 87.7% with close homologs in NCBI, therefore it can be recognized as a separate species (nucleotide similarity less than 95%) however it can be placed in same genus as *Escherichia* virus 186 as nucleotide homology was more than 70%.

BIS20 has highest similarity with *Salmonella enterica* sub species *enterica* serovar *Senftenberg* prophage (SEStP) sequence located in chromosome 1 (Acc. No; LS483465.1, sequence range 675,000–706,000) (Supplementary Table S5, Fig. 4), with only 87.7% nucleotide similarity. BLAST n results indicated that the proximal 21,368 bp of BIS20 has highest homology with SEStP (96%). This region has ORFs encoding for phage structural proteins. Remaining ~ 10 kb genome of BIS20 nucleotide sequence (22,078–29,477 bp) have varying degree of nucleotide homology (95–77%) with SEStP sequence whose ORFs encoded integration cassette, replication, and portal proteins among various other. Nucleotide sequence of two ORFs (26 and 35 Supplementary Table S4) ~ 710 bp each was missing from SEStP. ORF 26 encoded for phage repressor CI-C and ORF35 encoded a prophage / hypothetical protein having homology with prophage protein of *Klebsiella pneumoniae* (Acc no; SVY30080.1).

Amino acid sequence of BIS20 ORFs has highest homology with SEStP as indicated by BLASTp, except for few proteins associated with base plate assembly and prophage induction. Overall BIS20 phage proteins are conserved and exhibit a mosaic pattern as they are found in different prophages of *Salmonella sp.* and *Citrobacter freundii* (Supplementary Table S4). SEStP is integrated exclusively between Siderophore interacting protein and mismatch specific DNA glycosylase of *Salmonella* chromosome. Interestingly just before SEStP in chromosome 1 (Acc. No LS483465.1) there is another prophage integrated in the same position which has no similarity with BIS20. The genome and protein similarities were also profound for *Citrobacter freundii* prophage (Gen. Bank Acc. No CP026550.1), with 67.7% nucleotide similarity (Figs. 4, 5). Integration site in *Citrobacter sp.* prophage matches, is between *OmpA* family protein and DNA binding transcription regulator. While these prophages retained the structural module of *Enterobacteriaceae* phage P2, they have acquired different proteins for interacting with specific hosts. The sequence of *Salmonella* phage BIS20 is deposited in NCBI with Acc. No MZ520833. Comparison of BIS20 genome sequence with these prophages is given in Figs. 4, 6.

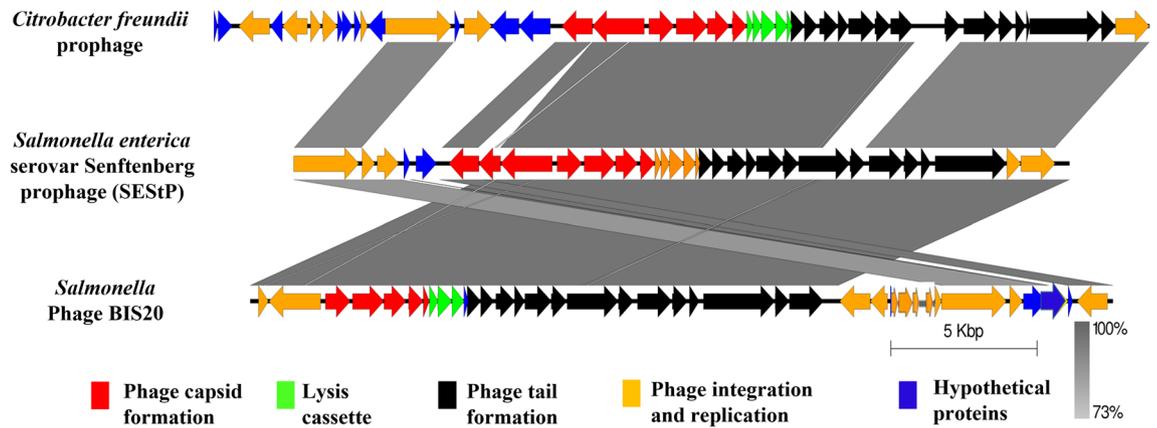


Figure 4. Genome comparison of BIS20 with closest prophage homologs *Salmonella enterica* serovar *Senftenberg* prophage (SEStP) and *Citrobacter freundii* prophage sequence. Various ORFs are color coded according to their putative function. (Scheme provided in Fig. 3).

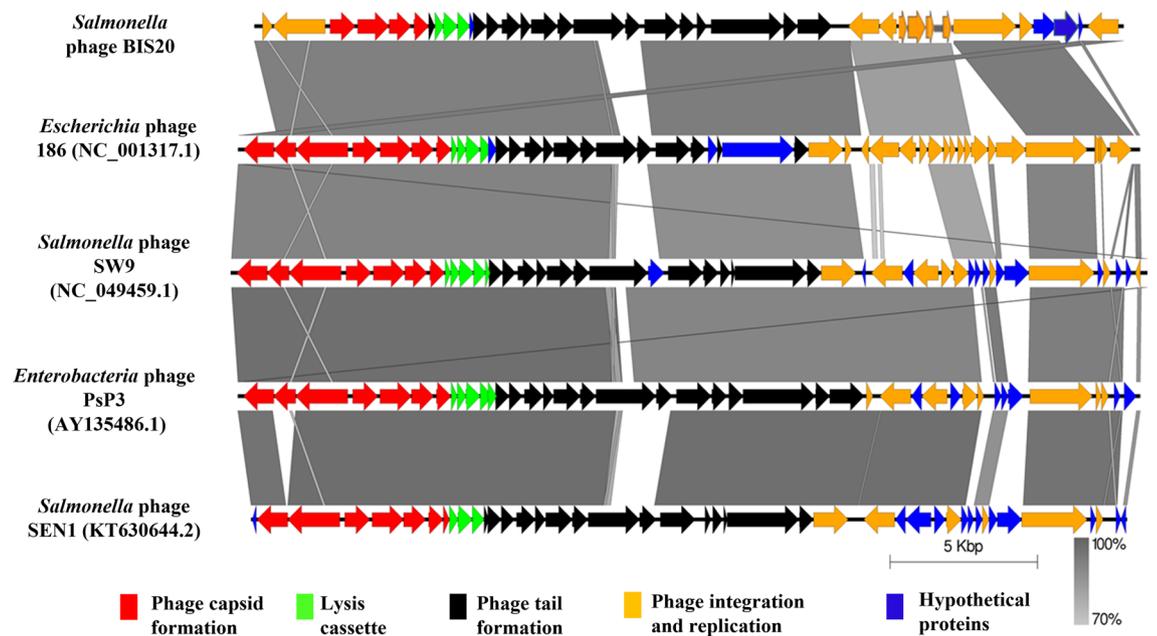


Figure 5. Genome comparison of BIS20 with closest bacteriophage homologs of genus *Eganvirus*. Various ORFs are color coded according to their putative function. (Scheme provided in Fig. 3).

BIS20 similarity with other bacteriophages. The closest phage homologs of BIS20 were *Escherichia* phage 186 (Phage 186) (82.4%), *Salmonella* phage SW9 (SW9) (71.2%), *Enterobacteria* phage PsP3 (PsP3) (69.7%) and *Salmonella* phage SEN1 (SEN1) (69.5%) as indicated by Easyfig and VIRIDIC analysis (Figs. 5, 6). These phages belong to genus *Eganvirus* of subfamily *Peduvirinae* (Supplementary Table S5, Figs. 5, 6). Interestingly the homology of all five phages had a conserved pattern. Genome of these phages was arranged in similar modules as BIS20. All of them exhibited 90–92% homology with BIS20 for the first 11 kb of genomic data having structural ORFs involved in capsid formation, lysis of host cell and part of tail formation. These ORFs were conserved and shared high sequence homology with each other and with BIS20. Followed by this region a ~1500 bp sequence of BIS20 encoding tail fiber domain and phage tail assembly protein (ORF17 and 18) was different in all phages and shared no sequence homology (Fig. 5). Interestingly, the initial 300 amino acids of tail fiber formation gene (TF) (ORF 17) were partially conserved in all phages (Supp. Figure S1) whereas remaining sequence has no homology. Amino acid sequence of TF from all phages was analyzed by GARD. It indicated presence of seven recombination hotspots of which one hotspot at amino acid 300 position was the most likely recombination site identified by eight breakpoint analysis (Fig. 7A). From this analysis it is presumed that active recombination at this site has enabled these phages to adapt to different hosts and expand their host range (Fig. 7B). It signifies the evolutionary contribution of such recombination events in giving rise to new phage species. We presume that one such event may have resulted in origin of BIS20 phage however it requires further experimental evaluation.

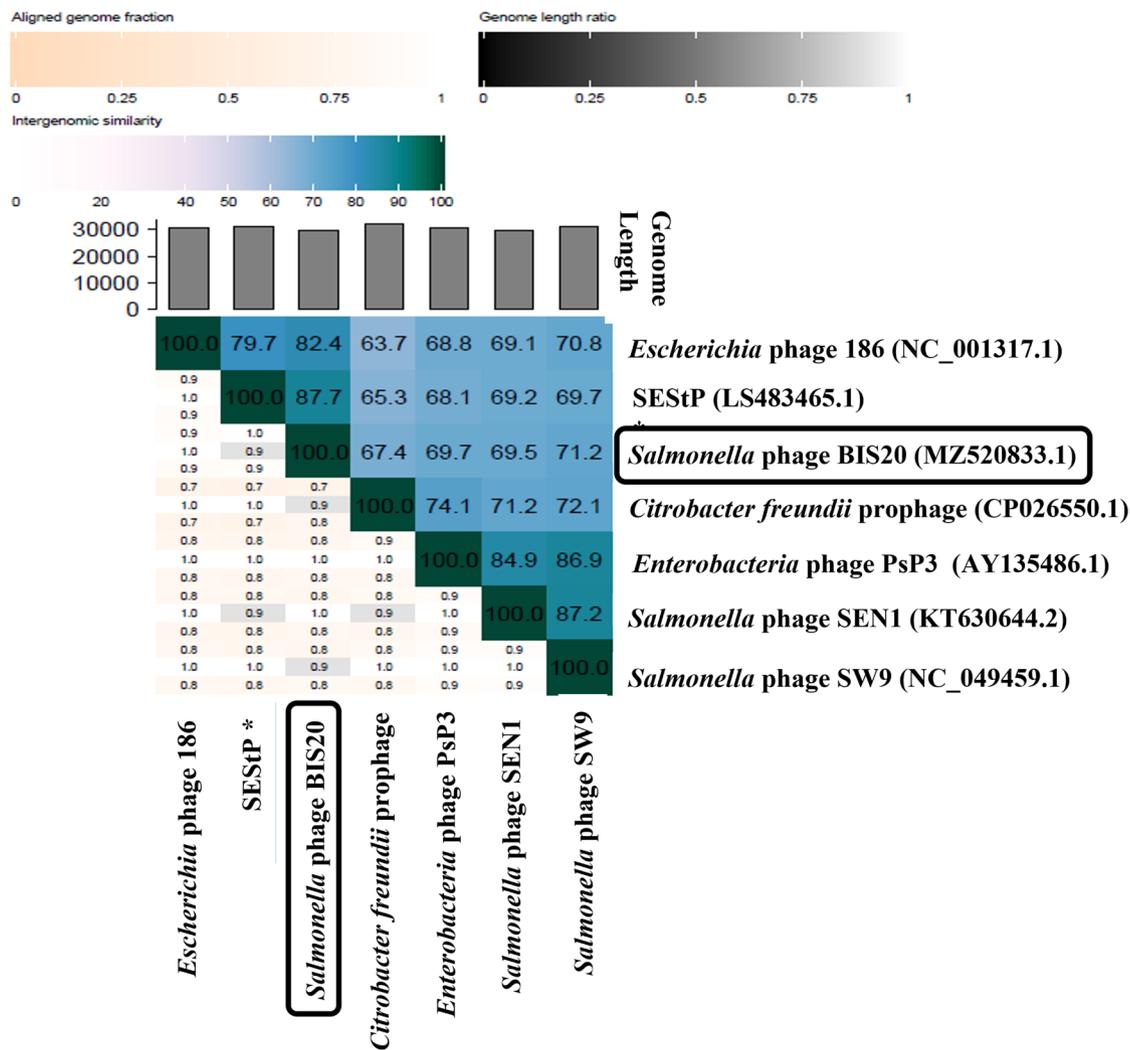


Figure 6. Heatmap generated by comparative genome analysis of *Salmonella* phage BIS20 with closest phage and prophage homologs in BLASTn using VIRIDIC software (<http://rhea.icbm.uni-oldenburg.de/VIRIDIC/>). Numbers in chart represent the homology percentage. *SEStP (*Salmonella enterica* sub species *enterica* serovar *Senftenberg* prophage).

Further BIS20 nucleotide sequence of roughly 3500/4000 bp (Fig. 5) encoding tyrosine recombinase XerC and *CI* repressor (lytic to lysogenic switch) was absent from three phages SW9, SEN1 and PsP3. BIS20 integration cassette had similar arrangement to phage 186 but with less amino acid conservations. It consists of an integrase, a transcription regulator protein *CI* (whose expression leads to lysogenization, an *Api* protein (similar to lambda COX) a putative excisionase whose expression promotes lytic mode of replication and an additional protein *CII* that positively regulates lysogeny (Supplementary Tables S4 and S5). These BIS20 proteins had (98%, 62.44% (No nucleotide homology), 70.11% (No nucleotide homology), and 85.88% amino acid homology with Phage 186 respectively.

Another section of BIS20 genome approximately ~795 bp and 2151 bp was missing from *Escherichia* phage 186 and other three phages (SEN1, PsP3 and SW9) respectively (Fig. 5). This sequence encoded four ORFs in BIS20 (DNA damage inducible protein (ORF33), two hypothetical proteins (ORF 34 and 36) and a putative prophage protein (ORF35). In phage 186 only two ORFs 35 and 36 were missing whereas all four ORFs were missing from remaining three phages.

A phylogenetic tree was constructed for BIS20 using its close phage and prophage homologs in BLAST n. The modified alignment after removing the gaps was 18,096 nucleotides long. Best fit model on the data was found to be GTR+I. Maximum Likelihood tree was produced by the TreeDyn program. Two major clusters were found in the tree; *Salmonella* phages were found in both the clusters. BIS20 was found to be closest with *Salmonella enterica* serovar *Senftenberg* prophage genome. These two genomes were sister to *Citrobacter freundii* prophage genome. Basal to this cluster was *Escherichia* phage 186. The other cluster contained three *Salmonella* (SW9, SEN1, S122) phages and *Enterobacter* phage PsP3. *Salmonella* phage SW9 and S122 were identical in sequence in this region. According to this analysis BIS20 is classified as a separate species in genus *Eganvirus* (Fig. 8).

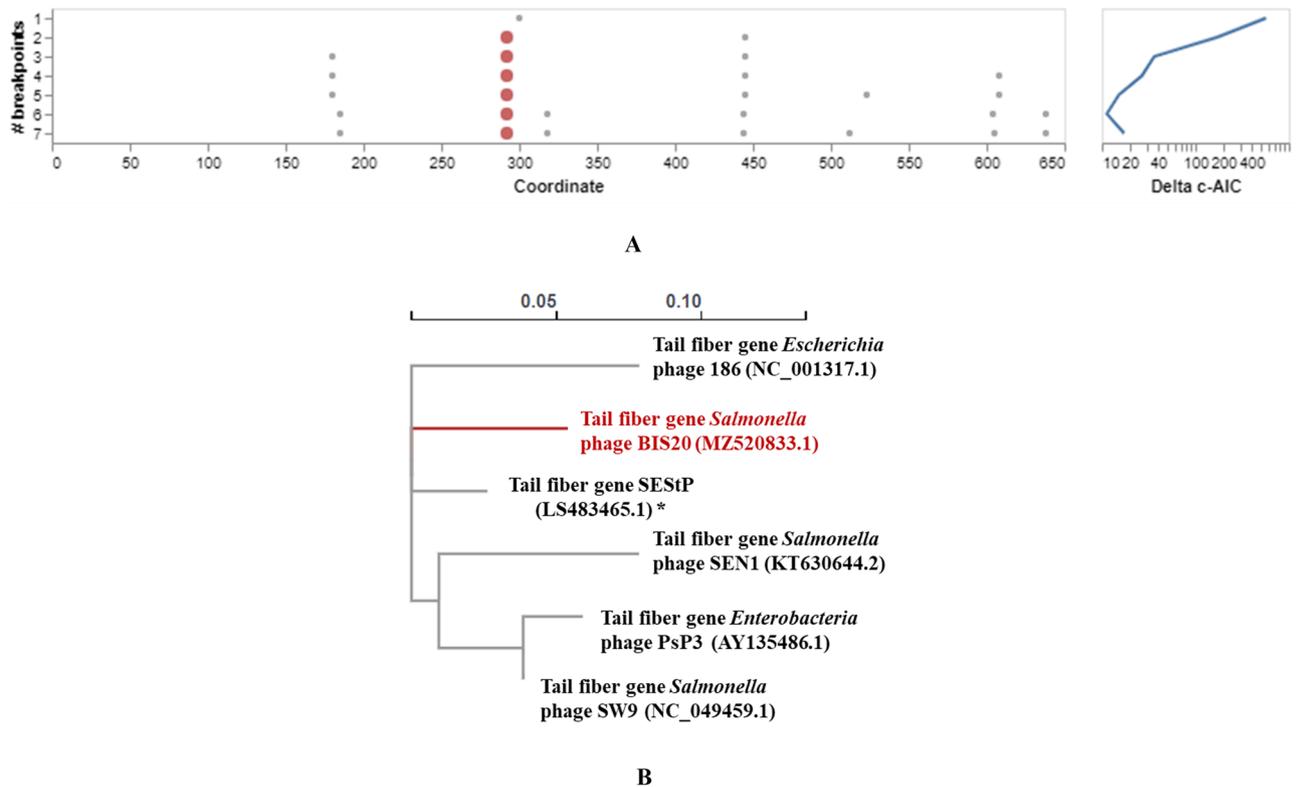


Figure 7. GARD analysis of the amino acid sequence of tail fiber gene with close homologs of BIS20. (A) Analysis identifies 8 breakpoints out of which the breakpoint at 300 position is the most probable breakpoint (represented with red dot). (B) Software generated phylogenetic tree made at coordinate range of 186–292 places BIS20 as a separate lineage.

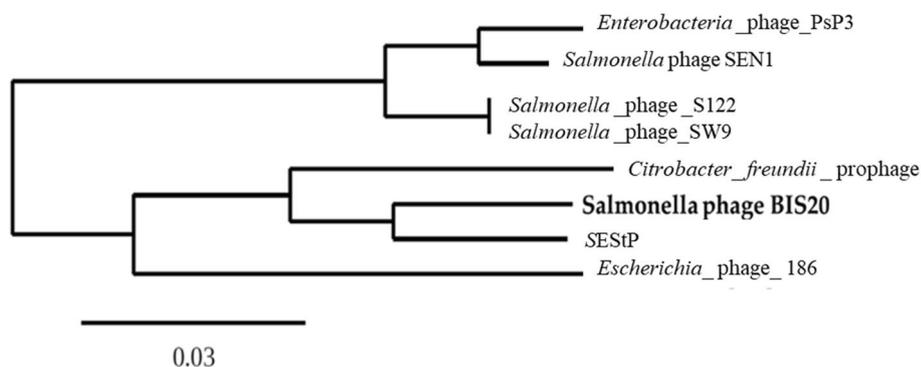


Figure 8. Phylogenetic tree of *Salmonella* phage BIS20 using whole-genome sequence of prophage and phage homologs in BLASTn. The maximum likelihood tree was generated using TreeDyn program and it provides a rough estimate of the relationships between related sequences in NCBI and *Salmonella* phage BIS20.

Discussion

S. Typhimurium is associated with recent outbreaks of gastroenteritis worldwide due to consumption of partially cooked or raw poultry meat^{1,2,23}. Temperate phages play a great role in host adaptability and pathogenicity. SE-BS17 used as host strain for BIS20 was characterized as *Salmonella enterica* subspecies *enterica* serovar Typhimurium by 16S rRNA sequencing. Although *S. Typhimurium* and *S. Typhi* are closely related to each other however BIS20 only infected *S. Typhimurium*, and not *S. Typhi*, this may indicate the recognition of only one host receptor by BIS20. Sustenance of only biological range of temperature and pH indicates that phage BIS20 may have adapted to its hosts' physiology over the course of evolution²⁴. *Salmonella enterica* harbors the richest pool of phages and prophages in its genome (roughly 9000 phage types)²⁵ that have been extensively studied^{6,26,27}. This study provides a detailed account of BIS20 physical properties and genome characterization. Nucleotide BLAST and VIRIDIC software analysis indicated that BIS20 is a new species of temperate phage closely related to members of genus *Eganvirus* of subfamily *Peduvirinae* family *Myoviridae*, this is in agreement with the fact

that majority *Salmonella* temperate phages discovered so far; belong to family *Myoviridae*^{28,29}. BIS20 has maximum homology with *Escherichia* phage 186 placed in the same genus and few prophage sequences in NCBI. The genome comprises of 53% GC content, well in alignment with its host strain (53.2%) indicating its evolutionary association with the host³⁰. Presence of tRNAs within the phage genome is known to broaden host range by codon usage compensation, however, no tRNAs were identified in BIS20 genome³¹, moreover no DNA or RNA polymerases were present in genome indicating its dependence upon host.

The genome of BIS20 exhibited mosaicism as different ORFs were related to several different *Salmonella enterica* and *Citrobacter freundii* prophage sequences in the NCBI. The same phenomenon is indicated in a detailed comparative analysis of *Enterobacteriaceae* lambdoid phages previously^{32–34}. Morphogenesis region is usually conserved in phages so was true for BIS20 whereas other genes have a mosaic arrangement probably due to previous crossovers and recombination events between phages. In BIS20 the genes responsible for binding with peptidoglycan; phage tail X protein, tail fiber and base plate assembly were less conserved and exhibited genetic diversity from both close prophage homologs (SEStP, *Citrobacter Sp.*).

BIS20 differs from its phage homologs by presence of unique long tail fiber gene (ORF17), tail fiber assembly genes (ORF18), as well as lytic to lysogenic switch proteins (ORF26–28). Interestingly these genes are different in all five members of *Eganvirus* genus. It is reported previously³⁵ that over the course of evolution tail fiber genes of apparently unrelated phages can exchange segments resulting in phage mosaic genomes. This can lead to a change in host range of bacteriophages as well. A study published by Sandmeyer³⁶ indicated presence of DNA invertase and recombination site like sequences at both ends of tail fiber gene segments conferring host specificity. They also reported that 5' end sequence of large tail fiber gene was conserved however 3' end conferring host specificity is different among different phages. When we compared BIS20 and *Escherichia* phage tail fiber gene (ORF18); first 789 bp at 5' end (~300 amino acids) have high homology (74%) however later 963 bp had ~49% nucleotide homology. Same was true for small tail fiber gene (ORF18, 49%). This observation indicates presence of a recombination hot spot in this region of tail fiber gene (ORF17) and same was found true by GARD analysis. Long tail fiber genes of three *Salmonella* specific bacteriophages SW9, S122 and SEN1 were also different from BIS20 although all of them infected *Salmonella enterica* serovars. The tail fiber gene of BIS20 was similar to the prophage homolog SEStP only. It is evolutionarily useful for phages to modify their proteins for interacting with host surface by site specific recombination and inversions. It suggests that BIS20 may have acquired these genes by similar mechanisms, GARD analysis indicated presence of 8 putative breakpoints in tail fiber genes of these phages out of which the best fit single breakpoint was at 300 amino acid position (Supp. Fig. S1). These two ORFs were present in *Salmonella* Senftenberg prophage (SEStP) sequence with 93 and 96% homology. BIS20 may have originated by recombination events between different phages and prophages from ancestral phage sequence by gene swapping^{9,37}.

BIS20 has a tyrosine recombinase; a site-specific integrase for lytic to lysogenic switch which is like that found in SEStP and *Escherichia* Phage 186. However, the recombinases of other three phages SW9, PsP3 and SEN1 shared 90% amino acid homology with each other but only share ~53% homology with *Escherichia* phage 186 and BIS20. Phage 186 has a complicated integration cassette where instead of only *Int*, *C* (repressor) and *cox* (excisionase lytic to lysogenic switch) protein of p2-phages; it has *Int*, *CI* (immunity repressor), *apI* (Excisionase) and *CII* (regulation) proteins³⁸. Both *CI* and *apI* genes of BIS20 have no nucleotide similarity with phage 186 however, protein sequence was 62% and 70% similar respectively. The exact mechanism of origin of this mosaicism is still debated. It is not clear whether it is product of random illegitimate recombination or due to specific recombination sites using DNA invertases and recombinases.

Temperate phages usually do not affect the ability of host to grow however BIS20 infection slowed down SE-BS17 growth to significant levels. Since the integration site for closest homolog SEStP is immediately before siderophore interacting protein, it may imply that integration event may have led to inactivation/ low efficiency of the siderophore interacting protein. Since this protein is crucial for iron uptake in the form of siderophores that play a critical role in bacterial growth³⁹ its inactivation may have resulted in poor growth of the *Salmonella* Typhimurium strain SE-BS17. This ability of BIS20 may be utilized to explore its role in host pathogenesis. Our findings have identified BIS20 as a new phage species of genus *Eganvirus* that is an important link between *Salmonella* prophages and *Escherichia* phage 186. Further biological analysis may shed light on its evolutionary significance.

Methods

Bacterial strain isolation and characterization. In this study we used a *Salmonella enterica* serovar Typhimurium isolate of poultry origin for phage isolation. This isolate was previously isolated and characterized in Molecular Virology Labs (MVL) CUI, naming SE-BS17 (Gene Bank Acc No. MZ503545)⁴⁰. Briefly sixty-five random poultry organ samples (Caecum, intestine, liver, and spleen) were collected from fourteen poultry farms in federal area supplying meat to retail poultry shops in Rawalpindi and Islamabad. The poultry samples were collected in sterile bags. Twenty-five grams of each poultry tissue sample (All tissue types pooled together) was homogenized to a fine paste using sterile surgical blades and pre enriched with 225 ml of 0.1% buffered peptone water (Oxoid). All samples were incubated overnight at 37 °C. After pre-enrichment 100 µl of broth culture was transferred to Selenite cystine broth (Oxoid) followed by incubation at 37 °C for 24 h. One loopful of enriched broths was streak plated onto Xylose Lysine Deoxycholate (XLD agar, Oxoid; CM0469) and incubated at 37 °C for 24 h. The plates were examined for typical colonies of *Salmonella*⁴¹. All isolates exhibiting typical *Salmonella* colonies on XLD plate were sub-cultured for biochemical characterization with API 20E strips (bio-Merieux; Ref No. 20100) as per manufacturers protocols.

Isolates identified as *Salmonella* (38 out of 65) by biochemical/numeric profile of API kit were further subjected to PCR amplification of 284 bp *InvA* gene for detection of *Salmonella* genus⁴² and *IroB* gene (606 bp)

for the detection of subspecies (Supplementary Table S1). Out of these isolates one strain SE-BS17 was used for bacteriophage isolation. Serovar of SE-BS17 was confirmed by amplification and sequencing of 16S rRNA gene (Macrogen, South Korea) (Supplementary Table S1). The strain SE-BS17 was identified as *Salmonella* Typhimurium. Further it was tested for antibiotic sensitivity against common antibiotics used in poultry by Kirby-Bauer disc diffusion method⁴³. SE-BS17 was tested for presence of beta lactamase (*bla*TEM-1, 643 bp)⁴⁴, extended spectrum beta lactamase (*bla*CTX-M, 754 bp)⁴⁵ and trimethoprim resistance gene *dfrA1* (474 bp)⁴⁶ by PCR amplification. The PCR was carried out as per methods given in publications cited; briefly reaction was carried out in total 40 µl adding 8 µl master mix (Solis Biodyne, Cat No. 04-11-00S15) 29 µl sterile water 2 µl (10 pmol) each forward and reverse primers and 1 µl of template DNA (250 ng). Amplification was carried out in thermo-cycler (Thermo fisher) with following conditions; initial denaturation at 95 °C for 5 min; 35 cycles of denaturation at 95 °C for 30 s; different annealing temperatures as per GC content of each primer given in respective publications. Elongation was carried out at 72 °C for 45 s.; and final extension at 72 °C for 8 min. PCR products were subjected to gel electrophoresis with DNA ladder (Solis Biodyne, Cat No.07-12-0000S) as molecular marker (Supplementary Table S1).

Bacteriophage isolation. Thirty-eight poultry samples (Caecum, intestine, liver, and spleen) previously tested positive for presence of different serovars of *Salmonella* as illustrated in “Bacteriophage isolation and characterization” were selected for phage isolation. The phage isolation was carried out as per protocols described by Duc⁴⁷. Briefly 1 g of each of four tissue types in one sample were taken in a Petri plate and washed twice with Phosphate Buffered Saline (PBS, pH 7.2) to remove debris. These tissues were then triturated into fine mince by scalpel blade and mixed in 10 ml buffered peptone water and incubated at 37 °C for 24 h. Next day the media was centrifuged at 8000 rpm for 10 min to remove bacterial debris and supernatant was filtered through 0.22 µm syringe filters (CNW technologies). After filtration 100 µl overnight culture of SE-BS17 was mixed in filtrate supplemented with 10 mM CaCl₂ and incubated at 37 °C for 24 h. Next day it was again centrifuged at 8000 rpm to remove bacterial cells and supernatant was filtered to obtain phage lysate. Phage lysate was stored at 4 °C until further use⁴⁸. Next, phage lysate was serially diluted (1:10) in LB broth. One drop of each dilution was spotted on an LB agar plate carrying 200 µl of SE-BS17 overnight culture in soft agar (Agar overlay method). All samples that produced individual plaques or lysis zones on SE-BS17 *Salmonella* strain after 24 h of incubation at 37 °C were considered positive and selected for further purification.

Plaque purification and phage stock preparation. Double layer agar overlay method was used for further plaque purification of *Salmonella* phage BIS20 (referred to as BIS20). One plaque from the phage-agar overlay plate was pierced using a tip-cut standard blue tip and the agar plug was ejected in 1 ml PBS (pH 7.2). The tubes were agitated for 1 h and then centrifuged at 10,000 rpm at 4 °C for 10 min. Supernatant was filtered through 0.22 µm syringe filters and kept at 4 °C until further use. The process was repeated twice to minimize the risk of contamination. For large scale preparation and amplification of BIS20; 1 L of LB medium (Oxoid) was inoculated (1: 10) with overnight grown culture of SE-BS17 and incubated at 37 °C with shaking till log phase (OD₆₀₀ 0.55) and then infected with phage at Multiplicity of Infection (MOI) 1. Infected cultures were incubated overnight as described previously⁴⁹. After overnight incubation, the culture was spun at 6,000 rpm for 10 min at 4 °C. Supernatant was filtered through 0.22 µm filter and stored at 4 °C. Phages were further concentrated by Polyethylene Glycol 8000 (PEG) (Sigma Aldrich Cat. No; 1546605) precipitation as reported elsewhere⁵⁰. Briefly, 10% w/v PEG was added to 1 L phage lysate and was kept without shaking overnight at 4 °C. Samples were centrifuged at 8,000 rpm at 4 °C (centrifuge; Helmer, Germany) for 40 min. Supernatant was decanted and the precipitate was dissolved in 2 ml Phosphate Buffered Saline (PBS) pH7.2 (137 mM NaCl, 2.7 mM KCl, 8 mM Na₂HPO₄, and 2 mM KH₂PO). The PEG-Enriched lysate was filtered through a 0.22 µm filter (HYSNY25022, Nylon syringe filters) and stored at 4 °C. The final phage titer was determined by the agar overlay spot method of serial two-fold dilutions of phage lysate as described in “Genome characterization”.

Phage host range. Host range of BIS20 was determined by its ability to produce clear zones of lysis on each strain by spot method (10 µl drops of BIS20 stock, 1.8×10^5 pfu/ml) using soft agar overlays⁵¹. A list of *Salmonella enterica* and other isolates used for host range determination is given in Supplementary Table S2. The host range was determined as per protocol used by Mazzocco. Briefly, 200 µl of overnight culture of each strain was mixed with 2.5 ml of soft agar (0.5% agar dissolved in LB) in 5 ml Kimax glass tubes (Sigma Aldrich Cat. No; Z255122). The mixture was then poured on to LB agar (Oxoid) plates and allowed to solidify for 20 min. Filtered phage lysate drops of BIS20 having titer 1.8×10^5 pfu/ml (10 µl) were placed on top of the seeded soft agar plates without dilutions. Presence of clear / turbid lysis zones was examined after 24 h. Bacterial strains were categorized as sensitive (+; lysis) or resistant (–; no-lysis) based on presence or absence of clear/ turbid lysis zones.

One step growth curve. Burst size and latent period of BIS20 were determined by the one step growth curve method with some modification to the method previously reported⁵². Briefly, 40 ml LB broth was inoculated with *Salmonella* strain SE-BS17 overnight culture (1:100) and incubated at 37 °C with shaking until the culture reached log phase (O.D.₆₀₀ 0.5) (2×10^7 CFU/ml). At log phase the culture was infected with BIS20 at MOI 1 (3.23×10^5 pfu/ml). The culture was left stationary for 15 min at 37 °C to allow phage adsorption and then centrifuged at 11,000 rpm for 1 min. The supernatant was transferred into new tubes and later titrated for enumeration of un-adsorbed phage particles. Cell pellet was washed three times with 1 ml fresh LB medium. After three washes the pellet was suspended in 30 ml fresh LB and allowed to grow at 37 °C for 1 h. One ml aliquots were collected from growing BIS20 infected SE-BS17 culture at 5 min interval for 60 min. The aliquots

were immediately centrifuged, filtered, and then titrated after dilutions using agar overlay method. Burst size was calculated using a formula given in Supplementary Table S3¹⁹.

Phage stability assay. The stability of BIS20 was tested by incubating fixed number of phages (1.8×10^5 pfu/ml) at different temperatures and pH values as described previously⁵³. Briefly, for assessment of thermal stability, 1.8×10^5 pfu of BIS20 were suspended in 1 ml PBS in triplicates and incubated at 37 °C, 60 °C, 80 °C and 100 °C respectively for 1 h. To assess BIS20 ability to sustain different pH values PBS with different pH values (2, 5, 7, 9, and 12) was prepared. pH was adjusted by either 6 N HCl or 6 M NaOH. Phages were suspended into this buffer to a final concentration of 1.8×10^5 pfu. These pH solutions were incubated at 37 °C for 1 h and then surviving phages were calculated immediately by double layer agar overlay plating method as described previously. The results are presented as average of triplicates.

Bacterial growth reduction assay. To determine the ability of BIS20 to lyse *Salmonella* Typhimurium (SE-BS17), bacterial growth reduction was quantified as described previously⁵⁴. Briefly, an overnight culture of SE-BS17 was diluted in 40 ml LB broth in four 100 ml glass flasks to a final concentration of 1.8×10^5 CFU/ml. Three flasks were inoculated with BIS20 at MOI 1. Fourth flask was used as bacteria only control with no phage added. After every 2 h, one ml culture was taken out of each flask using sterile disposable plastic pipette and its optical density (OD₆₀₀) was measured using a spectrophotometer (Hinodek, 721-100G). A standard curve was plotted to determine bacterial colony forming units at various optical density values by dilution plating. This curve was used to determine the CFU/ml at corresponding optical density at 600 nm for growth reduction assay. Results are presented as average of triplicates.

Electron microscopy. For negative staining, 5 µL of PEG purified phage sample ($\sim 6 \times 10^6$ pfu/ml) was applied onto glow-discharged carbon coated grids and incubated for five minutes at room temperature. Grids were washed with 5 µL of deionized water before incubating for 2 min with a 1% uranyl acetate solution. Electron micrographs were taken using a Phillips CM-10 microscope at the Manawatu Microscopy and Imaging Centre (MMIC, Massey University, Palmerston North, New Zealand)⁵⁵.

Sequence and bioinformatic analysis of BIS20. Bacteriophage DNA was isolated using phage DNA isolation kit of Norgen biotek Corp. (Cat # 46800) as per manufacturer's protocol⁵⁶. Phage genome sequencing was performed using the Illumina MiSeq platform at Massey Genome Service (Massey University, Palmerston North, New Zealand). Briefly, the DNA library was prepared using the Illumina Nextera™ XT library preparation kit_V2 (Illumina, San Diego, CA, USA). Phage genomic DNA was enzymatically sheared into random fragments. Barcoded Illumina adapters were added onto each end of the fragments during enrichment PCR. Libraries were QC checked with a PerkinElmer GX Touch HT Instrument using the labchip DNA high sensitivity and a Quant-iT dsDNA HS assay using a PerkinElmer Victor Plate Reader. Libraries were pooled before loading onto the Illumina MiSeq™ run. Libraries were sequenced using Illumina MiSeq Micro 300 cycle kit V1 (Illumina, San Diego, CA, USA) and generated 2×150 base paired end reads. PhiX control reads, and adapter sequences were removed using FASTQMC^{37,38}. Sequence reads then trimmed at an error probability of 0.01 (Pared score of Q20). Sequence data was tested for quality control (QC) using Fast QC (Version 0.11.5 released), Fast Q screen and Solexa QA + + (37). Contigs were assembled using Artemis software⁵⁹. The assembled genome was annotated using GAMOLA version 2, RAST (The server is freely available at <http://RAST.nmpdr.org>) and GeneMark.hmm (<http://opal.biology.gatech.edu/GeneMark/gmhmm.cgi>) software. ORFs and putative genes were identified by HHpred (available at <https://toolkit.tuebingen.mpg.de/tools/hhpred>) and RAST. Individual ORFs protein sequence data from BIS20 was analyzed by NACBI, BLASTp as well as HHpred. software. Phage is a linear dsDNA phage as no contigs assembled across both ends of the phage genome however the sequence of phage termini was analyzed using Phage Term software. It requires Individual BIS20 FASTAQ reads and a FASTA file of assembled genome. The PhageTerm software is freely available at pasture web server (<https://galaxy.pasteur.fr/>)²². The report generated by this software is included in Supplementary material file (Supplementary Fig. S2). The genome sequence of BIS20 is added to NCBI database with name “*Salmonella* phage BIS20” as per ICTV recommendations of naming a phage⁶⁰ under Acc. No. MZ520833.

Phylogenetic analysis. For genome wide comparison of BIS20 with related phage and prophage sequences VIRIDIC software was used. VIRIDIC applies the traditional algorithm recommended by International Committee on Taxonomy of Viruses (ICTV), subcommittee of Bacterial and Archaeal Viruses (BAVS) to calculate phage intergenomic similarities. VIRIDIC software has highest agreement with BLASTn. Genome comparison figures were generated using Easyfig software which is freely available (under a GPL license) for download at <https://mjsull.github.io/Easyfig/>⁶¹.

Since a simplified phylogenetic tree cannot adequately explain evolution of homologous sequences that are governed by gene conversion and recombination, we used multiple sequence alignment of tail fiber genes to search recombination breakpoints with high degree of quantitative support for their existence⁶². Recombination hotspots were identified by analysis through GARD software which is freely available at Datamonkey adaptive evolution server (<https://www.datamonkey.org/gard>).

To construct a traditional phylogenetic tree seven phage/prophage sequences having sequence homology with BIS20 were downloaded from NCBI. These sequences were aligned with BIS20 genome using MAFFT alignment⁶³ implemented in Geneious (v.8.1.9)⁶⁴. Alignment was trimmed from both ends and the gapped sequences / insertion and deletions were visually located and removed from the alignment. The alignments in Phylip format were imported in IQtree⁶⁵ online version is available at <http://iqtree.cibiv.univie.ac.at/> to reconstruct a Maximum

Likelihood tree based on the best fit model as implemented in the IQTree⁶⁶. IQTree implements Model Finder⁶⁷ to calculate best fit model on the data. The trees in Newick format were refined in the online TreeDyn tool⁶⁸ available at http://www.phylogeny.fr/one_task.cgi?task_type=treedyn and downloaded in pdf format.

Statistical analysis. Statistical analyses were performed with Origin 2019 (Origin Lab, Northampton, MA). The statistical significance was determined using two sample t test for specific comparisons. Statistical significance was reached at $p < 0.05$.

Data availability

The genome sequence of *Salmonella* phage BIS20 with the same name is deposited in NCBI GenBank with Acc. No MZ520833. Data is publicly available now.

Received: 3 November 2021; Accepted: 28 March 2022

Published online: 06 April 2022

References

- Arnedo Pena, A., Bellido Blasco, J. B., Romeu-Garcia, M., & Meseguer-Ferrer, N. Detection of foodborne *Salmonella* Typhimurium outbreaks. *Enferm. Infect. Microbiol.* 470–471 (2017).
- Xiang, Y. *et al.* Investigation of a salmonellosis outbreak caused by multidrug Resistant *Salmonella* typhimurium in China. *Front. Microbiol.* 11, 801. <https://doi.org/10.3389/fmicb.2020.00801> (2020).
- Hugas, M. & Beloelil, P. Controlling *Salmonella* along the food chain in the European Union—progress over the last ten years. *Eur. Commun. Dis. Bull.* 19(19), 1. <https://doi.org/10.2807/1560-7917.es2014.19.19.20804> (2014).
- Hayashi, T. *et al.* Complete genome sequence of enterohemorrhagic *Escherichia coli* O157: H7 and genomic comparison with a laboratory strain K-12. *DNA Res.* 8(1), 11–22 (2001).
- Hacker, J. & Carniel, E. Ecological fitness, genomic islands and bacterial pathogenicity. *EMBO Rep.* 2(5), 376–381 (2001).
- Mottawea, W. *et al.* *Salmonella enterica* prophage sequence profiles reflect genome diversity and can be used for high discrimination subtyping. *Front. Microbiol.* 9, 836. <https://doi.org/10.3389/fmicb.2018.00836> (2018).
- Kintz, E. *et al.* A BTP1 prophage gene present in invasive non-typhoidal *Salmonella* determines composition and length of the O-antigen of the lipopolysaccharide. *Mol. Microbiol.* 96(2), 263–275. <https://doi.org/10.1111/mmi.12933> (2015).
- Ashton, P. M. *et al.* Public health surveillance in the UK revolutionises our understanding of the invasive *Salmonella* Typhimurium epidemic in Africa. *Genome Med.* 9(1), 92. <https://doi.org/10.1186/s13073-017-0480-7> (2017).
- Hendrix, R. W., Smith, M. C., Burns, R. N., Ford, M. E. & Hatfull, G. F. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc. Natl. Acad. Sci. U.S.A.* 96(5), 2192–2197. <https://doi.org/10.1073/pnas.96.5.2192> (1999).
- Botstein, D. A theory of modular evolution for bacteriophages. *Ann. N. Y. Acad. Sci.* 354(1), 484–491 (1980).
- Van Dessel, W. *et al.* Complete genomic nucleotide sequence and analysis of the temperate bacteriophage VWB. *Virology* 331(2), 325–337 (2005).
- Lucchini, S., Desiere, F. & Brüssow, H. Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory. *J. Virol.* 73(10), 8647–8656 (1999).
- Kutter, E. *et al.* Evolution of T4-related phages. *Virus Genes* 11(2–3), 285–297 (1995).
- Chibani-Chennoufi, S., Canchaya, C., Bruttin, A. & Brüssow, H. Comparative genomics of the T4-Like *Escherichia coli* phage JS98: Implications for the evolution of T4 phages. *J. Bacteriol.* 186(24), 8276–8286 (2004).
- Haggård-Ljungquist, E., Halling, C. & Calendar, R. DNA sequences of the tail fiber genes of bacteriophage P2: Evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. *J. Bacteriol.* 174(5), 1462–1477 (1992).
- Brüssow, H., Canchaya, C. & Hardt, W.-D. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol. Mol. Biol. Rev. MMBR* 68(3), 560–602. <https://doi.org/10.1128/MMBR.68.3.560-602.2004> (2004).
- Panis, G., Franche, N., Méjean, V. & Ansaldo, M. Insights into the functions of a prophage recombination directionality factor. *Viruses* 4(11), 2417–2431. <https://doi.org/10.3390/v4112417> (2012).
- De Paepe, M. *et al.* Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: The role of Rad52-like recombinases. *PLoS Genet.* 10(3), 1–15. <https://doi.org/10.1371/journal.pgen.1004181> (2014).
- Adams, M. H. *Bacteriophages*. CiteSeer (1959).
- Hyman, P. Phages for phage therapy: Isolation, characterization, and host range breadth. *Pharmaceuticals* 12(1), 1. <https://doi.org/10.3390/ph12010035> (2019).
- Lwoff, A. Lysogeny. *Bacteriol. Rev.* 17(4), 269–337. <https://doi.org/10.1128/br.17.4.269-337.1953> (1953).
- Garneau, J. R., Depardieu, F., Fortier, L.-C., Bikard, D. & Monot, M. PhageTerm: A tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. *Sci. Rep.* 7(1), 8292. <https://doi.org/10.1038/s41598-017-07910-5> (2017).
- Ford, L. *et al.* Seven *Salmonella* Typhimurium outbreaks in Australia linked by trace-back and whole genome sequencing. *Foodborne Pathog. Dis.* 15(5), 285–292 (2018).
- Jończyk, E., Klak, M., Miedzybrodzki, R. & Górski, A. The influence of external factors on bacteriophages-review. *Folia Microbiol.* <https://doi.org/10.1007/s12223-011-0039-8> (2011).
- Wahl, A., Battesti, A. & Ansaldo, M. Prophages in *Salmonella enterica*: A driving force in reshaping the genome and physiology of their bacterial host?. *Mol. Microbiol.* 111(2), 303–316. <https://doi.org/10.1111/mmi.14167> (2019).
- Gao, R. *et al.* Comparative genomic analysis of 142 bacteriophages infecting *Salmonella enterica* subsp. *enterica*. *BMC Genomics* 21, 1–13 (2020).
- Moreno Switt, A. I. *et al.* *Salmonella* phages and prophages: Genomics, taxonomy, and applied aspects. in *Salmonella: Methods and Protocols: Second Edition* (2014).
- Lavigne, R., Seto, D., Mahadevan, P., Ackermann, H.-W. & Kropinski, A. M. Unifying classical and molecular taxonomic classification: Analysis of the Podoviridae using BLASTP-based tools. *Res. Microbiol.* 159(5), 406–414 (2008).
- Ackermann, H.-W. Bacteriophage observations and evolution. *Res. Microbiol.* 154(4), 245–251. [https://doi.org/10.1016/S0923-2508\(03\)00067-6](https://doi.org/10.1016/S0923-2508(03)00067-6) (2003).
- Lu, N. *et al.* Characterization and genome analysis of the temperate bacteriophage φSAJS1 from *Streptomyces avermitilis*. *Virus Res.* 265, 34–42. <https://doi.org/10.1016/j.virusres.2019.03.006> (2019).
- Hyman, P. & Abedon, S. T. Bacteriophage host range and bacterial resistance. *Adv. Appl. Microbiol.* 1, 1. [https://doi.org/10.1016/S0065-2164\(10\)70007-1](https://doi.org/10.1016/S0065-2164(10)70007-1) (2010).
- Pedulla, M. L. *et al.* Origins of highly mosaic mycobacteriophage genomes. *Cell* 113(2), 171–182. [https://doi.org/10.1016/s0092-8674\(03\)00233-2](https://doi.org/10.1016/s0092-8674(03)00233-2) (2003).

33. Juhala, R. J. *et al.* Genomic sequences of bacteriophages HK97 and HK022: Pervasive genetic mosaicism in the lambdoid bacteriophages. *J. Mol. Biol.* **299**(1), 27–51. <https://doi.org/10.1006/jmbi.2000.3729> (2000).
34. Proux, C. *et al.* The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like Siphoviridae in lactic acid bacteria. *J. Bacteriol.* **184**(21), 6026–6036. <https://doi.org/10.1128/JB.184.21.6026-6036.2002> (2002).
35. Sandmeier, H. Acquisition and rearrangement of sequence motifs in the evolution of bacteriophage tail fibres. *Mol. Microbiol.* **12**(3), 343–350. <https://doi.org/10.1111/j.1365-2958.1994.tb01023.x> (1994).
36. Sandmeyer, H. Acquisition and rearrangement of sequence motifs in the evolution of bacteriophage tail fibres. *Mol. Microbiol.* **12**(3), 343–350 (1994).
37. Hatfull, G. F. & Hendrix, R. W. Bacteriophages and their genomes. *Curr. Opin. Virol.* **1**(4), 298–303. <https://doi.org/10.1016/j.coviro.2011.06.009> (2011).
38. Christie, G. E. & Calendar, R. Bacteriophage P2. *Bacteriophage* **6**(1), e1145782. <https://doi.org/10.1080/21597081.2016.1145782> (2016).
39. Behnsen, J. & Raffatelli, M. Siderophores: More than Stealing Iron. *MBio* **7**(6), 16. <https://doi.org/10.1128/mBio.01906-16> (2016).
40. Sattar, S. *et al.* Genome analysis and therapeutic evaluation of a novel lytic bacteriophage of *Salmonella* Typhimurium: Suggestive of a new genus in the subfamily Vequintavirinae. *Viruses* **14**(2), 241 (2022).
41. Budiyati, T. *et al.* Prevalence, antibiotic resistance and plasmid profiling of *Salmonella* in catfish (*Clarias gariepinus*) and tilapia (*Tilapia mossambica*) obtained from wet markets and ponds in Malaysia. *Aquaculture* **372–375**, 127–132. <https://doi.org/10.1016/j.aquaculture.2012.11.003> (2013).
42. Shanmugasamy, M., Velayutham, T. & Rajeswar, J. Inv A gene specific PCR for detection of *Salmonella* from broilers. *Vet. World* **4**(12), 562 (2011).
43. Biemer, J. J. Antimicrobial susceptibility testing by the Kirby-Bauer disc diffusion method. *Ann. Clin. Lab. Sci.* **3**(2), 135–140 (1973).
44. Frye, J. G. & Jackson, C. R. Genetic mechanisms of antimicrobial resistance identified in *Salmonella enterica*, *Escherichia coli*, and *Enterococcus* spp. isolated from U.S. food animals. *Front. Microbiol.* **4**, 1–22. <https://doi.org/10.3389/fmicb.2013.00135> (2013).
45. Ramachandran, A., Shanthi, M. & Sekar, U. Detection of blaCTX-Mextended spectrum betalactamase producing *Salmonella enterica* serotype typhi in a tertiary care centre. *J. Clin. Diagn. Res.* **11**(9), 21–24. <https://doi.org/10.7860/JCDR/2017/30150.10637> (2017).
46. El-Tayeb, M. A., Ibrahim, A. S. S., Al-Salamah, A. A., Almaary, K. S. & Elbadawi, Y. B. Prevalence, serotyping and antimicrobials resistance mechanism of *Salmonella enterica* isolated from clinical and environmental samples in Saudi Arabia. *Braz. J. Microbiol.* **48**(3), 499–508. <https://doi.org/10.1016/j.bjm.2016.09.021> (2017).
47. Duc, H. M., Son, H. M., Honjoh, K. & Miyamoto, T. Isolation and application of bacteriophages to reduce *Salmonella* contamination in raw chicken meat. *LWT* **91**, 353–360. <https://doi.org/10.1016/j.lwt.2018.01.072> (2018).
48. Bao, H., Zhang, H. & Wang, R. Isolation and characterization of bacteriophages of *Salmonella enterica* serovar Pullorum. *Poult. Sci.* <https://doi.org/10.3382/ps.2011-01496> (2011).
49. Carey-Smith, G. V., Billington, C., Cornelius, A. J., Hudson, J. A. & Heinemann, J. A. Isolation and characterization of bacteriophages infecting *Salmonella* spp. *FEMS Microbiol. Lett.* **1**, 1. <https://doi.org/10.1111/j.1574-6968.2006.00217.x> (2006).
50. Yamamoto, K. R., Alberts, B. M., Benzinger, R., Lawhorne, L. & Treiber, G. Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large-scale virus purification. *Virology* **40**(3), 734–744. [https://doi.org/10.1016/0042-6822\(70\)90218-7](https://doi.org/10.1016/0042-6822(70)90218-7) (1970).
51. Mazzocco A., Waddell T.E., Lingohr E. Enumeration of bacteriophages using the small drop plaque assay system. In Clokie M.R., Kropinski A.M. (eds) *Bacteriophages. Methods in Molecular Biology*, vol. 501 (2009).
52. Nagayoshi, Y. *et al.* Physiological properties and genome structure of the hyperthermophilic filamentous phage ϕ OH3 which infects thermus thermophilus HB8. *Front. Microbiol.* **7**, 50. <https://doi.org/10.3389/fmicb.2016.00050> (2016).
53. Capra, M. L., Quiberoni, A. & Reinheimer, J. Phages of *Lactobacillus casei/paracasei*: Response to environmental factors and interaction with collection and commercial strains. *J. Appl. Microbiol.* <https://doi.org/10.1111/j.1365-2672.2005.02767.x> (2006).
54. O'Flynn, G., Coffey, A., Fitzgerald, G. F. & Ross, R. P. The newly isolated lytic bacteriophages st104a and st104b are highly virulent against *Salmonella enterica*. *J. Appl. Microbiol.* **101**(1), 251–259. <https://doi.org/10.1111/j.1365-2672.2005.02792.x> (2006).
55. Mikalová, L. *et al.* Novel temperate phages of *Salmonella enterica* subsp. *salamae* and subsp. *diarizonae* and their activity against Pathogenic *S. enterica* subsp. *enterica* Isolates. *PLoS ONE* **12**(1), 34. <https://doi.org/10.1371/journal.pone.0170734> (2017).
56. Pereira, C. *et al.* Bacteriophages with potential to inactivate *Salmonella* Typhimurium: Use of single phage suspensions and phage cocktails. *Virus Res.* **220**, 179–192. <https://doi.org/10.1016/j.virusres.2016.04.020> (2016).
57. Aronesty, E. ea-utils: 'Command-line tools for processing biological sequencing data'. <http://code.google.com/p/ea-utils>.
58. Aronesty, E. Comparison of sequencing utility programs. <https://doi.org/10.2174/1875036201307010001> (2013).
59. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics (Oxford, England)* **27**(4), 578–579. <https://doi.org/10.1093/bioinformatics/btq683> (2011).
60. Adriaenssens, E. M. & Rodney Brister, J. How to name and classify your phage: An informal guide. *Viruses* **1**, 1. <https://doi.org/10.3390/v9040070> (2017).
61. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics (Oxford, England)* **27**(7), 1009–1010. <https://doi.org/10.1093/bioinformatics/btr039> (2011).
62. Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. & Frost, S. D. W. GARD: A genetic algorithm for recombination detection. *Bioinformatics* **22**(24), 3096–3098. <https://doi.org/10.1093/bioinformatics/btl474> (2006).
63. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**(2), 511–518. <https://doi.org/10.1093/nar/gki198> (2005).
64. Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)* **28**(12), 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199> (2012).
65. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**(1), 268–274. <https://doi.org/10.1093/molbev/msu300> (2014).
66. Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**(W1), W232–W235. <https://doi.org/10.1093/nar/gkw256> (2016).
67. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermini, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**(6), 587–589. <https://doi.org/10.1038/nmeth.4285> (2017).
68. Chevenet, F., Brun, C., Bañuls, A.-L., Jacq, B. & Christen, R. TreeDyn: Towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **7**(1), 439. <https://doi.org/10.1186/1471-2105-7-439> (2006).

Author contributions

S.S., conceptualization, experimental plan, project design, funding acquisition, bioinformatic and software analysis, manuscript preparation. I.U., B.S., A.P., and F.S. planned and executed experiments (IU with highest contribution). S.K., sequencing support, E.A., M.B. genome data analysis. T.S. statistical analysis, I.A. phylogenetic analysis, K.I. and N.A.S. data curation, N.B., S.J. manuscript revision, A.G. provided logistical and technical support.

Funding

This research was funded by the Higher Education Commission (HEC) Pakistan, NRPU vide grant number 8091.

Competing interests

Author Ibrar Ahmed is employed at Alpha genomics whereas Sofia Khanum, Marc Bailie and Eric Altermann are employed at AgResearch. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. Authors declare no conflict of interest.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-09733-5>.

Correspondence and requests for materials should be addressed to S.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022